PRELIMINARY EXPERIMENT FOR KOREA DATA CUBE WITH KOMPSAT-3/3A ARD BASED ON OPEN DATA CUBE PLATFORM

Kwangseob Kim[1], and Kiwon Lee*[2]

[1]Assistant Professor, Department of Computer Software, Kyungmin University
545, Seobu-ro, Uijeongbu-si, Gyeonggi-do, 11618, Korea (ROK)
Email: kwangseob@kyungmin.ac.kr

[2]Professor, Department of Applied Convergence Security, Hansung University
Samseon-dong, Seongbuk-gu, Seoul, 02876, Korea (ROK)
Email: kilee@hansung.ac.kr

KEY WORDS: Korea Data Cube, KOMPSAT, Open Data Cube, Analysis Ready Data

ABSTRACT: Analysis Ready Data (ARD) is an interoperable product with other data in remote sensing applications, allowing immediate analysis. Committee on Earth Observation Satellites (CEOS) developed and released the guideline for specifications of ARD, as CEOS ARD for Land (CARD4L) for surface reflectance, surface temperature, and synthetic aperture radar (SAR)-based normalized back scatter. US Geological Survey (USGS) has provided Landsat ARD products though collection-2 since 2020. Sentinel ARD products, which complied with the CARD4L specification, have also been generated. ARD products support time series analysis. Therefore, a large amount of data must be processed and archived. Storage is a crucial factor. Amazon Simple Storage Service (S3) is one of the cloud-based storage services, a technology called object storage. It can store most data formats, including unstructured types of data, with quickly accessible functionality through an Application Programming Interface (API). Archived ARD products can be linked with Earth Engine (GEE), providing about 70 petabytes of ready-to-use data. Another component for ARD applications is Open Data Cube (ODC), an open-source platform. In most cases, this platform has been used to country-wide application. This study presents the preliminary experiment results to develop or build the Korea Data Cube (KDC) using image data sets of the KOMPSAT series. In this study, ARD is dealt with the surface reflectance of KOMPSAT 3/3A. The cloud computing environment is also used with OpenStack, an open-source cloud platform. It is expected that this approach will be the core scheme for future KDC, as well as prospective data cube of other countries on the planning stage or initial developing stage.

1. INTROCUCTION

Since the 2020s, various technology development and system construction projects have been underway to use the accumulated satellite imagery data effectively. These technology developments directly relate to cloud computing technology (Guo et al., 2020; Sudmanns et al., 2022; Xu et al., 2022). These trends are also influenced by the growing popularity of commercial cloud services such as Amazon Web Services (AWS), Google Cloud, and Microsoft Azure. Meanwhile, the demand for platforms that can store and manage large amounts of satellite imagery data for private use for public purposes has increased. In response to this demand, the GEO has supported Open Data Cube, developed in the mid-2010s, as a core common platform.

Open Data Cube (ODC) is an open-source software project that provides a unified interface for accessing, managing, and analyzing large-scale Earth observation (EO) data. It is a big data solution that works with massive datasets, often terabytes or petabytes. ODC achieves this using a distributed architecture, meaning the data is stored and processed across multiple machines. This allows ODC to scale to vast datasets without compromising performance. ODC also provides many features specifically designed for working with big data. Meanwhile, the data collected from these ODCs are for ARD.

ARD is a term that refers to satellite data products that have been processed and formatted to a certain standard so that they can be easily used for analysis by different users and applications. ARD products can reduce the time and effort required for data preparation and increase the consistency and quality of the analysis results. ARD products benefit EO applications, such as monitoring land use, climate change, natural disasters, and biodiversity. In this regard, CEOS is developing basic guidelines and publishing standard specifications for creating ARD data. In particular, surface reflectance and surface temperature from optical satellite imagery are recognized as essential data for terrestrial applications, so they have been specially prepared for CARD4L. CARD4L data is available from various sources, including Copernicus Open Access Hub, Sentinel Hub, USGS Earth Explorer, and Planet Explorer. US Geological Survey (USGS) has provided Landsat ARD products through collection-2 (Dwyer et al., 2018). Sentinel ARD products, which complied with the CARD4L specification, have also been generated. ARD products support time series analysis. As of 2023, there are over 30 countries that have built or are using ODC. Some countries with operational ODCs include Australia, Colombia, Switzerland, and Brazil. Most countries are building ODCs using ARDs based on Landsat or Sentinel-2 imagery, but some countries also apply ARDs for their satellite imagery to ODCs (Santoro et al., 2023).

Currently, Korea is developing an integrated system for the production of KOMPSAT ARD (Kang and Yoo, 2019) and is considering various platforms to provide it. This study presents the preliminary experiment results to develop or build the Korea Data Cube (KDC) using image data sets of the KOMPSAT series.

2. BASIC SCHEME FOR EXPERIMENTS

The experimental study conducted in this research requires several components to integrate a service system based on ODC. Currently, KOMPSAT imagery is not in the process of being certified by CARD4L (Figure 1), but it does meet CARD4L's basic specifications internally. The specifications of the KOMPSAT-3/3A imagery, which are the subject of this study, are presented in Table 1.

Figure 2 shows the workflow of this study. First, the request and order process for KOMPSAT-3/3A images is needed. KOMPSAT-3/3A data consists of Level 1 geometrically corrected bundled data image sets. Bundled data refers to data that is provided individually according to spectral bands. Second, the data acquisition process with ordered imagery is necessary. Third, the absolute atmospheric corrected product for KOMPSAT-3/3A at Level-2 is processed to generate KOMPSAT-3/3A ARD with calibration (Kim and Lee, 2021). These processes are carried out via cloud service as a software-as-a-service (SaaS) running on OpenStack. Fourth, the KOMPSAT-3/3A ARD product is registered and indexed in ODC. Next, KOMPSAT-3/3A ARD is ingested into ODC and registered in SaaS for users' applications.

In this study, we utilized OpenStack cloud instead of AWS or other commercial cloud services. OpenStack is an open-source project providing a flexible platform that can be employed to construct various satellite image processing systems. This can assist organizations in meeting their specific requirements. In simple comparisons between OpenStack and AWS, both OpenStack and AWS offer object storage services like Swift and S3, which can be utilized to store satellite images. These services are scalable and reliable, suitable for storing large volumes of data. Both OpenStack and AWS also provide computing services such as Nova and EC2 for running image processing applications. These services offer a range of instance types that can be customized to the application's specific needs. Furthermore, OpenStack and AWS offer analytics services such as EMR and Athena for analyzing satellite images. These services provide a variety of tools for extracting insights from the data. Table 2 shows service projects in OpenStack and the version applied in this study.

Figure 1. CEOS ARD progress (Oliver, 2023)

Table 1. Summary of KOMPSAT-3/3A

Mission characteristics	KOMPSAT-3	KOMPSAT-3A
Launch time	18 May 2012	26 Mar 2015
Orbit altitude	685 km	528 km
Swath width	≥ 15 km (at nadir)	≥ 12 km (at nadir)

Ground sample distance	Pan: 0.7 m MS: 2.8 m	Pan: 0.55 m MS: 2.2 m IR: 5.5 m
Spectral bands	Pan: 450-900 nm Blue: 450-520 nm Green: 520-600 nm Red: 630-690 nm NIR: 760-900 nm	Pan: 450-900 nm Blue: 450-520 nm Green: 520-600 nm Red: 630-690 nm NIR: 760-900 nm MWIR: 3300-5200 nm
Radiometric resolution	14 bits	14 bits

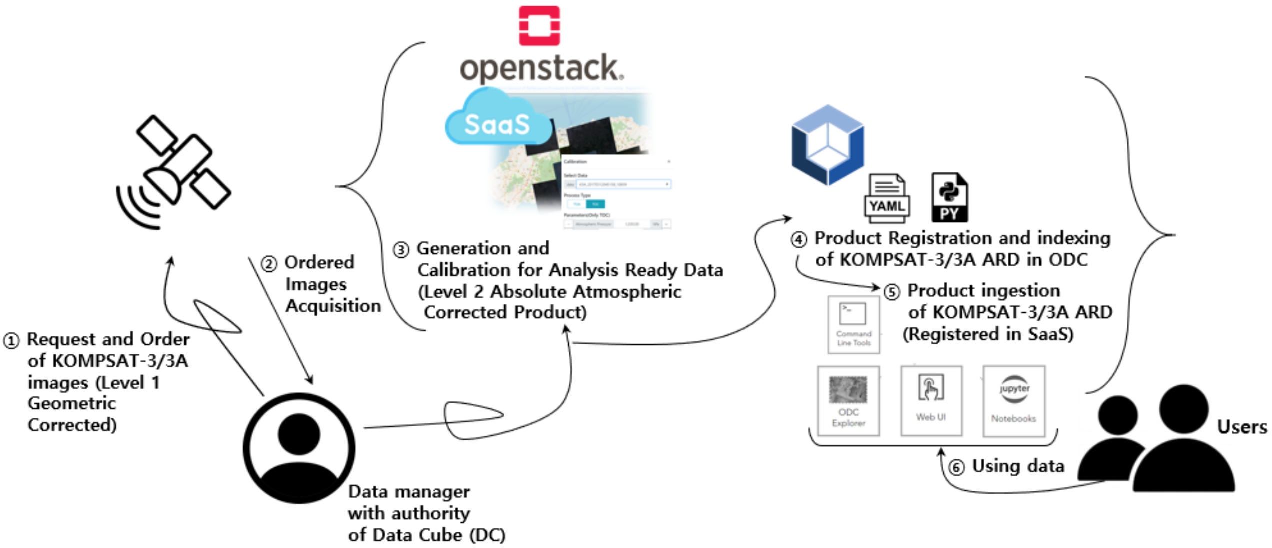

Figure 2. Workflow of this study.

Table 2. OpenStack specification applied in this study

Service type	Project name (Version)
Platform	OpenStack (Wallaby)
Identity service	Keystone (19.0.1)
Image service	Glance (22.0.1)
Networking service	Neutron (18.1.1)
Compute service	Nova (23.0.3)
Block storage service	Cinder (18.0.1)
Dashboard	Horizon (19.2.1)
Shared file systems service	Manila (12.0.1)

3. IMPLEMENTED RESULTS

Figure 3 shows the result of the ARD generation and ODC ingestion processes implemented in this study. This result corresponds to steps 3, 4, and 5 in Figure 2. This study selected Jeju Island as a case study area for a preliminary experiment. The data was purchased from KOMPSAT 3/3A Level 1 bundled image sets. A total of 13 were processed by surface reflectance (SR) ARD and applied to the ODC platform configured to register KOMPSAT. Figure 3 shows the results of uploading the data to a separate SaaS for ODC registration. The SaaS system can automatically visualize in the form of web mapping if KOMPSAT series Level 1 data is uploaded. Figures 3(a) and (b) show the uploading processes of KOMPSAT-3A bundled data. SaaS consists is comprised of a web mapping interface. Users can upload auxiliary files in Blue, Green, Red, Near Infrared, and Extensible Markup Language (XML), and the list of KOMPSAT 3/3A data registered and Figure 3(c) and (d) are the web mapping results. Figure 3(e) lists the results of processing Level 1 data by SR ARD. The SaaS includes atmospheric and surface reflectance processing functions for KOMPSAT-3/3A. This system has all the meta information about KOMPSAT data. It performs SR-generating processing in real-time by absolute atmospheric correction algorithms based on atmospheric pressure, water vapor, ozone amount, and aerosol optical parameters. The indexed results, including bundled and surface reflectance images (KOMPSAT3A_TOCR), are displayed on the ODC interface. Data cube-based data processing is possible by indexing and ingesting SR ARD to ODC after processing.

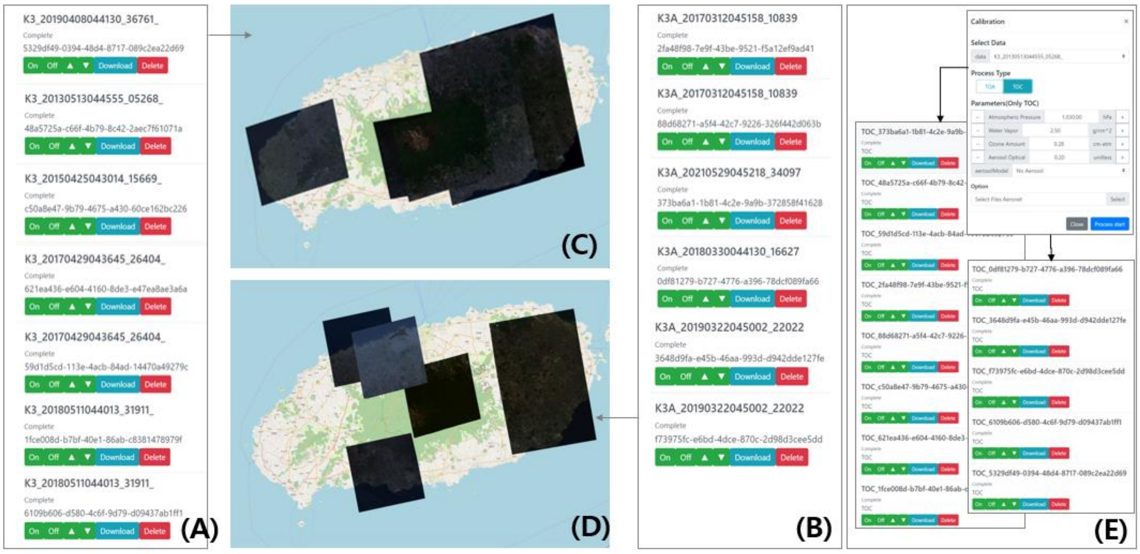

Figure 3. KOMPSAT-3/3A visualization and processing results on SaaS: (a) KOMPSAT-3 data list, (b) KOMPSAT-3A data list, (c) KOMPSAT-3 web mapping results, (d) KOMPSAT-3A web mapping results, and (e) KOMPSAT-3/3A spectral reflectance processing results.

Figure 4 shows the results of computed Normalized Difference Vegetation Index (NDVI) images through ODC, where KOMPSAT-3/3A products are registered. The visual processing results were processed through the ODC User Interface (UI), a sub-project of ODC, to view them on the web. Figure 4(a) shows the process of selecting the registered data. Users can choose the Level 1 data uploaded through the data visualization and processing SaaS and the processed SR data. At this time, the data is processed as time series data by being ingested into ODC. Figure 4(b) shows the results of selecting the processing area. Users can set the min/max latitude, longitude values, and period (start/end) through the geospatial bounds item. Latitude and longitude values can be set using the map drawing function. Figure 4 (c) and (d) show the processing results. The processing registered in ODC internally pre-sets the index colors for the range and visualizes the results based on this. The NDVI processing results are represented in blue, white, and green, respectively, from -1.0 to -0.1, 0.0, and 0.1 to 1.0. This can be modified at any time in the ODC UI.

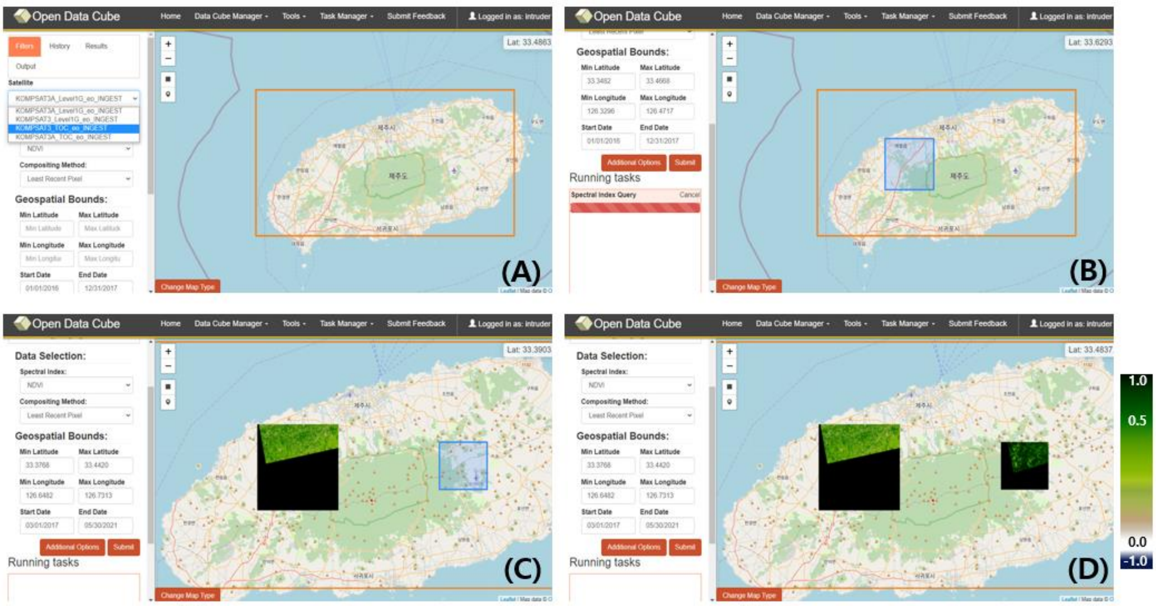

Figure 4. KOMPSAT-3/3A Jeju region NDVI processing request and visualization results: (a) Selection of ingested satellite data, (b) Request after selecting the area and period to be processed, (c) and (d) Web mapping results after processing.

4. CONCLUDING REMARKS AND PLANS

The motivation of this study is to generate temporary ARDs from KOMPSAT images and integrate them into ODCs. In addition, the study aims to run ODCs on OpenStack instead of AWS.

The preliminary experimental results of the application of the ODC platform, which is able to process KOMPSAT ARD products and perform time series processing simultaneously, not only serve as baseline data for future KDC operations, but also as practical examples for the generation of satellite information products and the development of time series monitoring systems. These results will be further refined through ongoing research.

In both the distribution and application areas of satellite imagery, ARD is expected to continue to evolve, resulting in a variety of user-friendly data. This advancement will lead to faster and more accurate results in earth monitoring, which is a primary goal of satellite imagery utilization. Korea plans to launch EO satellites for various applications. With the continuous operation of the KOMPSAT series, the Compact Advanced Satellite (CAS 500) satellite (Land Satellite No. 1) has been launched and is currently operational for land monitoring in Korea, and there are ongoing plans to launch more new satellites for forest applications and water resources monitoring. In order to ensure quick and easy data access for users, it is essential to conduct research and establish ARD production systems that comply with the CARD4L specifications for satellite image provision. It is also necessary to develop services that can be instantly processed using platforms and technologies such as ODC.

We hope that this research will proceed with plans that take into account the following points, if it is possible to the support of concerned organizations or sectors of civilization. First, it will provide baseline data for CARD4L approval of KOMPSAT-3/3A ARD. Second, it will develop an integrated operating system with AWS. This will be a multi-cloud system using OpenStack. Third, it will be extended to a system that can process backscatter ARD images of KOMPSAT-5 SAR images in addition to optical satellites. Fourth, it will add Cloud Optimized GeoTIFF (COG)-SpatioTemporal Asset Catalog (STAC) functionality to facilitate the user experience for web services. Fifth, it will be enhanced with functionality related to the ODC sandbox and GEE. Finally, it will be developed as a system for international data sharing and compatibility by enabling GEO-GEOSS or Open Earth Alliance (OEA) connectivity.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00241346).

References

Kang, C., and Yoo, D., 2019. KARI's Satellites & ARD, CEOS LSI-VC-8 Agenda Session 9, Anchorage, USA.

Kim, K., and Lee, K., 2021. An Implementation of Open Source-Based Software as a Service (SaaS) to Produce TOA and TOC Reflectance of High-Resolution KOMPSAT-3/3A Satellite Image, *Remote Sensing*, 13(22), 4550.

Dwyer, J. L., David, P. R., Brian, S., Calli, B. J., Hankui, K. Z., and Leo, L., 2018. Analysis Ready Data: Enabling Analysis of the Landsat Archive. *Remote Sensing*, 10(9), 1363.

Guo, H., Nativi, S., Liang, D., Craglia, M., Wang, L., Schade, S., Corban, C., He, G., Pesaresi, M., Li, J., Shirazi, Z., Liu, J., and Annoni, A., 2020. Big Earth Data science: an information framework for a sustainable planet. *International Journal of Digital Earth*, 13(7), pp.743-767.

Santoro, M., Mazzetti, P., and Nativi, S., 2023. Virtual earth cloud: a multi-cloud framework for enabling geosciences digital ecosystems. *International Journal of Digital Earth*, 16(1), pp. 43-65.

Oliver, S., 2023. An ARD Quality Framework to achieve interoperability and multi-mission data analysis, https://www.ard.zone/ard23

Sudmanns, M., Augustin, H., Killough, B., Giuliani, G., Tiede, D., Leith, A., Yuan, F., and Lewis, A., 2022. Think global, cube local: an Earth Observation Data Cube's contribution to the Digital Earth vision. *Big Earth Data*, 7(3), pp.831-859.

Xu, C., Du, X., Fan, X., Giuliani, G., Hu, Z., Wang, W., Liu, J., Wang, T., Yan, Z., Zhu, J., Jiang, T. and Huadong Guo, H., 2022. Cloud-based storage and computing for remote sensing big data: a technical review, *International Journal of Digital Earth*, 15(1), pp.1417-1445.